Лекция 3

Теоретико-информационна сигурност

3.1 Теория на Shannon

Ще въведем понятието съвършена (или безусловна) сигурност на криптосистема. Това понятие предполага опонент неограничен изчислителен ресурс. Очаквано, криптосистемите, които са сигурни при това допускане, трябва да изпълняват доста ограничителни условия. Както и в предните две глави, ще означаваме с $\mathcal{P}$, $\mathcal{C}$ и $\mathcal{K}$ съответно множествата на откритите текстове, ключовете и криптотекстовете. С всяко от тези множества свързваме случайна величина:

$$P = (\mathcal{P}, p_P), C = (\mathcal{C}, p_C), K = (\mathcal{K}, p_K),$$

където p_P, p_C, p_K са разпределенията съответно на P, C и K. Естествено е да приемем, че случайните величини P и K са независими. Това означава, че потребителят не взима под внимание откритият текст при избора на ключ. Означаваме с $\mathcal{C}(k)$ множеството на всички криптотекстове, които се получават при шифриране с ключ k.:

$$\mathcal{C}(k) = \{E_k(x) \mid x \in \mathcal{P}\}.$$

Ясно е, че разпределението на случайна величина C се определя от разпределенията на P и K. В сила е[1]

$$p(C = c) = \sum_{k \in \mathcal{K}} p(K = k) p(P = D_k(c)). \tag{3.1}$$

Да разгледаме криптосистема с

$$\mathcal{P} = \{a, b, c, d\}, \mathcal{C} = \{1, 2, 3\}, \mathcal{K} = \{k_1, k_2, k_3\}.$$

вероятностните разпределения p_P и p_K са

$$p(P = a) = \frac{1}{4}, p(P = b) = \frac{3}{10}, p(P = c) = \frac{3}{20}, p(P = d) = \frac{3}{10},$$

[1] По нататък ще пишем $p_{\mathcal{C}}(c)$ или $p(C = c)$.

$$p(K = k_1) = \frac{1}{4}, p(K = k_2) = \frac{1}{2}, p(K = k_3) = \frac{1}{4}.$$

Шифрирането и дешифрирането се задават със следната таблица:

	a	b	c	d
k_1	3	4	2	1
k_2	3	1	4	2
k_3	4	3	1	2

Сега с помощта на (3.1) получаваме

$$\begin{aligned}
p(C = 1) &= p(K = k_1)p(P = d) + p(K = k_2)p(P = b) + p(K = k_3)p(C = c) = 0.2625 \\
p(C = 2) &= p(K = k_1)p(P = c) + p(K = k_2)p(P = d) + p(K = k_3)p(C = d) = 0.2625 \\
p(C = 3) &= p(K = k_1)p(P = a) + p(K = k_2)p(P = a) + p(K = k_3)p(C = b) = 0.2625 \\
p(C = 4) &= p(K = k_1)p(P = a) + p(K = k_2)p(P = c) + p(K = k_3)p(C = a) = 0.2125
\end{aligned}$$

Криптотекстовете се оказват доста равномерно разпределени. Сега при зададены криптотекст $y \in \mathcal{C}$ и открит текст $x \in \mathcal{P}$ можем да изчислим условната вероятност $p(C = y|P = x)$ Това е вероятността да получим криптотекст y при положение, че е шифриран откритият текст x Тази вероятност се получава от равенството:

$$P(C = y|P = x) = \sum_{k:x=D_k(y)} p(K = k).$$

Тук сумата е по всички ключове, които дешифрират y в x. Така от даденото разпределение на K и от таблицата за шифриране/дешифриране получаваме:

$$\begin{array}{ll}
P(C = 1|P = a) = 0 & p(C = 1|P = b) = 0.5 \\
P(C = 2|P = a) = 0 & p(C = 2|P = b) = 0 \\
P(C = 3|P = a) = 0.75 & p(C = 3|P = b) = 0.25 \\
P(C = 4|P = a) = 0.25 & p(C = 4|P = b) = 0.25
\end{array}$$

$$\begin{array}{ll}
P(C = 1|P = c) = 0.25 & p(C = 1|P = b) = 0.25 \\
P(C = 2|P = c) = 0.25 & p(C = 2|P = b) = 0.75 \\
P(C = 3|P = c) = 0 & p(C = 3|P = b) = 0 \\
P(C = 4|P = c) = 0.5 & p(C = 4|P = b) = 0
\end{array}$$

Прикриптанализ на даден криптотекст y ние наблюдаваме случайната величина C. Възниква естественият въпрос, какъв е най-вероятният открит текст x при наблюдаван криптотекст y. така ние се интересуваме по-скоро от вероятността $p(P = x|C = y)$, която се получава от формулата на Bayes:

$$p(P = x|C = y) = \frac{p(P = x)p(C = y|X = x)}{P(C = y)}. \tag{3.2}$$

Тези условни вероятности лесно се получават от разпрделенията на случайните величини P и K и таблицата на ширираният/дешифрираният. така сега имаме:

$$\begin{array}{ll} P(P=a|C=1)=0 & p(P=a|C=2)=0 \\ P(P=b|C=1)=0.571 & p(P=b|C=2)=0.143 \\ P(P=c|C=1)=0.143 & p(P=c|C=2)=0 \\ P(P=d|C=1)=0.286 & p(P=d|C=2)=0.857 \end{array}$$

$$\begin{array}{ll} P(P=a|C=3)=0.714 & p(P=a|C=4)=0.294 \\ P(P=b|C=3)=0 & p(P=b|C=4)=0.353 \\ P(P=c|C=3)=0.286 & p(P=c|C=4)=0.353 \\ P(P=d|C=3)=0 & p(P=d|C=4)=0 \end{array}$$

Сравнявайки тези резултати с разпределението на P, получаваме информация за изпратения открит текст. Ако полученият криптотекст е 1, то е невъзможно да е бил изпратен откритият текст a; нещо повече – най-вероятният открит текст в този случай е b. При получен криптотекст 2 най-вероятният открит текст е d, а откритите текстове a и c са невъзможни и т.н. така евентуалният опонент, за който приемаме, че познава криптосистемата (т.е. разпределенията $p_P, p_K, p_{C|P}$ и $p_{P|C}$, както и таблицата за шифриране/дешифриране) получава някаква информация[2] за ткрития текст като наблюдава единствено криптотекста. От гледна точка на комуникиращите страни това трябва да бъде избягнато. Идеалната ситуация е криптотекстът да не дава никаква информация за открития текст при всеки избор на открит текст x и криптотекст y. Криптосистема с това свойство наричаме *съвършена* или *съвършено сигурна*.

Дефиниция 3.1. Казваме, че една криптосистема е *съвършена* (или *съвършено сигурна*), ако за всяко $x \in \mathcal{P}$ и всяко $y \in \mathcal{C}$ е в сила

$$p((P=x|C=y) = p(P=x).$$

С други думи опонентът не знае нищо повече за изпратеното съобщение от страничния наблюдател. От равенството на Bayes следва, че една криптосистема е съвършена тогава и само тогава, когато за всяко $x \in \mathcal{P}$ и всяко $y \in \mathcal{C}$ е изпълнено

$$p(C=y|P=x) = p(C=y).$$

Теорема 3.2. *За всяка съвършена криптосистема с множество от откри текстове $\mathcal{P}$, множество от криптотекстове $\mathcal{C}$ и множество от ключове $\mathcal{K}$ е изпълнено*

$$|\mathcal{K}| \geq |\mathcal{C}| \geq |\mathcal{K}|.$$

Доказателство. Тъй като шифриращата функция е инективна, имаме $|\mathcal{C}| \geq |\mathcal{P}|$. Ще приемем без ограничение на общността, че всички криптотекстове могат да бъдат получени, т.е. че $P(C=y) > 0$ за всяко $y \in \mathcal{C}$. В противен случай просто ще модифицираме $\mathcal{C}$, отхвърляйки невъзможните елементи. Сега за всяко съобщение $x \in \mathcal{P}$ и всеки криптотекст $y \in \mathcal{C}$ е в сила (от бележката преди теоремата):

$$p(C=y|P=x) = p(C=y) > 0.$$

[2]Понятието количество информация по Shannon дефинираме строго по-нататък.

Следователно за всяка двойка (x, y) съществува ключ k, за който $p(K = k) > 0$ и $x = D_k(y)$.

Фиксираме открития текст $x = x_0$. Сега на всеки криптотекст y можем да съпоставим ключ $k(y)$, за който $x_0 = D_{k(y)}(y)$. Разбира се, на различни криптотекстове ще съответстват различни ключове, откъдето $|\mathcal{K}| \geq |\mathcal{C}|$. □

Теорема 3.3. *(*C. Shannon*)* Нека $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ е криптосистема, за която

$$|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|.$$

Тази система е съвършена тогава и само тогава, когато

- използването на различните ключове е равновероятно, т.е.

$$p(K = k) = \frac{1}{|\mathcal{K}|}.$$

  за всяко $k \in \mathcal{K}$.

- за всяка двойка $x \in \mathcal{P}$, $y \in \mathcal{C}$ съществува единствен ключ k, за който $E_k(x) = y$.

Доказателство. 1) Да допуснем, че криптосистемата е съвършена. Тогава за всяка двойка $x \in \mathcal{P}$, $y \in \mathcal{C}$ съществува ключ $k \in \mathcal{K}$, за който $D_k(y) = x$. От определението за криптосистема следва, че $E_k(x) = y$. Оттук получаваме, че (поради $|\mathcal{C} = \mathcal{K}|$)

$$|\{E_k(x) \mid k \in \mathcal{K}\}| = |\mathcal{K}|,$$

т.е. не съществуват ключове k_1, k_2, $k_1 \neq k_2$, за които $E_{k_1}(x) = E_{k_2}(x) = y$. така за всички двойки (x, y) съществува единствен ключ $k \in \mathcal{K}$ със свойството $E_k(x) = y$. В частност за тройката $x, y, k)$ имаме

$$p(C = y|P = x) = p(K = k).$$

Ще докажем първата част, т.е. че използването на всички ключове е равновероятно. Нека $|\mathcal{K}| = n$ и $\mathcal{P} = \{x_i \mid i = 1, \ldots, n\}$. да фиксираме $y \in \mathcal{C}$ ид а номерираме ключовете така, че $E_{k_i}(x_i) = y$. Тъй като криптосистемата е съвършена $p(P = x_i|C = y) = p(P = x_i)$ и получаваме

$$p(P = x_i|C = y) = p(P = x_i) = \frac{p(C = y|P = x_i)p(P = x_i)}{p(C = y)} == \frac{p(K = k_i)p(P = x_i)}{p(C = y)}.$$

Следователно, $p(C = y) = p(K = k_i)$. От това равенство и от факта, че y е фиксирано, следва, че всички ключове се използват с една и съща вероятност

$$p(K = k) = p(C = y) = \frac{1}{|K|}.$$

2) Сега да предположим, че

- $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$,

- Използването на всеки ключ е равно вероятно,
- за всяка двойка $x \in \mathcal{P}$, $y \in \mathcal{C}$ съществува единствен ключ $k \in \mathcal{K}$, за които $E_k(x) = y$.

Ще докажем справедливостта на равенството $p(P = x|C = y) = p(P = x)$. От равновероятността на ключовете получаваме

$$\begin{aligned} p(C = y) &= \sum_{k \in \mathcal{K}} p(K = k) p(P = D_k(y)) \\ &= \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} p(P = D_k(y)) \end{aligned}$$

Освен това доколкото за всяка двойка $x \in \mathcal{P}$, $y \in \mathcal{C}$ съществува единствен ключ k, преобразуващ x в y имаме

$$\sum_{k \in \mathcal{K}} p(P = D_k(y)) = \sum_{x \in \mathcal{P}} p(P = x) = 1.$$

Следователно, $p(C = y) = \dfrac{1}{|\mathcal{K}|}$. Ако $y = E_k(x0$, то

$$p(C = y|P = x) = p(K = k) = \frac{1}{|\mathcal{K}|}.$$

Сега от теоремата на Bayes получаваме

$$p(P = x|C = y) = \frac{p(P = x)p(C = y)|P = x)}{p(C = y)} = \frac{p(P = x) \cdot \frac{1}{|\mathcal{K}|}}{\frac{1}{|\mathcal{K}|}} = p(P = x).$$

□

Не е трудно да се докаже, че шифърът на Vigenére, при който дължината на ключа еравна на дължината на открития текст и шифърът на Vernam са съвършено сигурни криптосистеми. По принцип това е един и същ шифър, трансформиращите операции на който се задават с таблицата на с
хард бирането (изваждането) на подходящо избран модул. В случая на Vigenére това е $\mathbb{Z}_{26}^m$, докато в случая на Vernam – $\mathbb{Z}_2^m$. Достатъчно е да приеме, че разпределението на ключовете е равномерно. Условията от теоремата на Shannon се проверяват тривиално.

3.2 Ентропия и взаимна информацията

Нека X е случайна величина, дефинирана върху множеството $\mathcal{X} = \{x_1, x_2, \ldots, x_m\}$ чрез $Pr_X\{X = x_i\} = p_i, 1 \leq i \leq m$. Мярка за количеството информация, която получаваме при сбъдване на събитието $X = x_i$ е

$$J(p_i) = -\log_2 Pr\{X = x_i\} = -\log_2 p_i. \tag{3.3}$$

Тук основата на логаритъма е несъществена; ако използваме основа 2, както в (3.3), то единицата за количество информация се нарича *бит*. С други думи, един бит е количеството информация, което получаваме от събитие появяващо се с вероятност $\frac{1}{2}$, въобще $J(\frac{1}{2^k}) = k$. Достоверното събитие, т.е. такова, което се появява с вероятност 1 не носи информация, а събитие, имащ вероятност близка до нула, носи безкрайно много информация.

Дефиниция 3.4. Математическото очакване за величината $J\left(Pr_X\{X = x\}\right)$ наричаме *ентропия* на X и означаваме с $H(X)$. Ентропия на случайна величина X с разпределение $\boldsymbol{p} = (p_1, \ldots, p_m)$ означаваме още с $H(\boldsymbol{p})$:

$$H(\boldsymbol{p}) = H(X) = E\left(J(Pr_X\{X = x\})\right) = \sum_{i=1}^{m} p_i J(p_i) = -\sum_{i=1}^{m} p_i \log_2 p_i.$$

Ентропията е функционал от разпределението на X. Тя не зависи от стойностите, които X приема, а единствено от техните вероятности. При пресмятането на горния израз считаме, че $0 \log 0 = 0$.[3] Така добавянето на елементи към $\mathcal{X}$, имащи нулева вероятност, не изменя ентропията. Да отбележим, че $H(X)$ не зависи от стойностите, които приема случайната величина X, а само от разпределението $\boldsymbol{p}$. Ентропията $H(X)$ на случайната величина X може да бъде интерпретирана по следния начин:

- като очакваното количество инфромация, получено при наблюдение на X;
- като мярка за нашата несигурност по отношение на X;
- като очаквания брой битове, необходими за описанието на възможните изходи на X.

Понятието ентропия може да бъде въведено и аксиоматично. Приемането на някои естествени свойства за аксиоми води с необходимост до логаритмична мярка за количество инфромация. Може да се докаже, че ако за редицата от симетрични функции $H_m(p_1, p_2, \ldots, p_m)$ са в сила свойствата:

(P1) $H_2(\frac{1}{2}, \frac{1}{2}) = 1$ (нормиране);

(P2) $H_2(p, 1-p)$ е непрекъсната функция на p;

(P3) $H(p_1, \ldots, p_n) = H(p_1, \ldots, p_{n-2}, p_{n-1} + p_n) + (p_{n-1} + p_n) H(\frac{p_{n-1}}{p_{n-1}+p_n}, \frac{p_n}{p_{n-1}+p_n})$;

то $H_m(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i$ за всяко $m = 2, 3, \ldots$. Съществуват и други аксиоматизации за понятието ентропия, които водят до същата функция [15].

В случая $n = 2$ ентропията $H(p, 1-p)$ означаваме с $h(p)$. Очевидно

$$h(p) = -p \log_2 p - (1-p) \log_2 1 - p, \quad 0 \leq p \leq 1. \tag{3.4}$$

Функцията $h(p)$ може да бъде доопределена в краищата на интервала $[0, 1]$ чрез $h(0) = h(1) = 0$. Графиката на $h(p)$ има следния вид:

[The graph of the function h(p)]

[3] Това може да се оправдае от съображения за непрекъснатост, тъй като $x \log x \to 0$ при $x \to 0$.

Пример 3.5. (хвърляне на монета) Да разгледаме експеримент, състоящ се в хвърлянето на монета. Пространството от елементарни събития е $\mathcal{X} = \{$лице,герб$\}$. Нека

$$Pr(\text{лице}) = p, \quad Pr(\text{герб}) = 1 - p.$$

Ентропията се задава с (3.4). Ако играта е честна, т.е. $p = 1 - p = \frac{1}{2}$, имаме $h(\frac{1}{2}) = 1$, което означава, че един бит е достатъчен за представянето на изхода от едно хвърляне. При "нечестна" игра ($p \neq 1 - p$) имаме $h(p) < 1$. Известно е, че можем да се доближим произволно близо до тази стойност, разглеждайки достатъчно дълги редици от опити (хвърляния). Нека например, $p = \frac{1}{4}$. Тогава $h(\frac{1}{4}) \approx 0.8113$. Да разгледаме две последователни хвърляния на монета. Възможните четири изхода са дадени по-долу заедно с тяхното представяне:

два последователни опита		вероятност	представяне
лице	лице	1/16	111
лице	герб	3/16	110
герб	лице	3/16	10
герб	герб	9/16	0

Очакваната дължина на представянето за единизход е

$$\frac{1}{2}(3 \cdot \frac{1}{16} + 3 \cdot \frac{3}{16} + 2 \cdot \frac{3}{16} + 1 \cdot \frac{9}{16}) = \frac{27}{32} \approx 0.843.$$

Ако групираме изходите на 3, 4 и т. н. броя експерименти, ще получим още по-добри приближения. Лесно се проверява, че всяка редица от нули и единици, получена от конкатенацията на низовете 111, 110, 10, 0 се разбива еднозначно на поднизове от този тип (вж също раздел **??**.

За всяка случайна величина X съществува описание с очаквана дължина между $H(X)$ и $H(X) + 1$.

Сега ще разширим дефиницията на понятието ентропия з адвойка случайни величини. Това не е принципно различна ситуация от разглеждана досега, защото (X, Y) може да се разглежда като случайна величина, приемаща векторни стойности. Нека X и Y са случайни величини със съвместно разпределение

$$p_{X,Y}(x, y) := Pr_{X,Y}\{X = x, Y = y\}.$$

Условната вероятност за $X = x$ при условие, че се е случило $Y = y$, означаваме с

$$p_{X|Y}(x|y) := Pr_{X|Y}\{X = x \mid Y = y\}.$$

Съвместната ентропия $H(X, Y)$ на двойка дискретни случайни величини (X, Y) със съвместно разпределение $p(x, y)$ дефинираме като

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y) \log p_{X,Y}(x, y),$$

или с други думи

$$H(X,Y) = -E \log p_{X,Y}(X,Y).$$

Ще дефинираме условната ентропия на една случайна величина X при зададена Y като очакваната стойност на ентропиите на условните разпределения, усреднена по случайната величина задаваща условието. Така нашата несигурност за X, при наблюдавано $Y = y$, се измерва от

$$H(X|Y=y) = -\sum_x p_{X|Y}(x|y) \cdot \log_2 p_{X|Y}(x|y). \tag{3.5}$$

Горният израз може да се интерпретира и като количеството информация, което очакваме да получим от наблюдението на случайната величина X, ако е известно, че е реализирано събитието $Y = y$.

Дефиниция 3.6. *Условна ентропия* (или *еквивокация* на X при известно Y е очакваната стойност на ентропията (3.5): $\sum_y p_y(y) H(X|Y=y)$.

Изразът за условната ентропия може да бъде преобразуван както следва:

$$\begin{aligned} H(X|Y) &= \sum_y p_Y(y) H(X|Y=y) \\ &= -\sum_y p_Y(y) \sum_x p_{X|Y}(x|y) \log_2 p_{X|Y}(x|y) \\ &= -\sum_x \sum_y p_Y(y) p_{X|Y}(x|y) \log_2 p_{X|Y}(x|y) \\ &= -\sum_x \sum_y p_{X,Y}(x,y) \cdot \log_2 p_{X|Y}(x|y). \end{aligned}$$

Последното равенство може да се използва за дефиниция на понятието условна ентропия.

Теорема 3.7. *(Chain rule)*

$$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X).$$

По-специално,

$$H(X,Y) \leq H(X) + H(Y)$$

като равенство се достига тогава и само тогава, когато X и Y са независисми случайни величини.

Доказателство. Преобразувайки формулата за съвместната ентропия на X и Y получаваме

$$\begin{aligned} H(X,Y) &= -\sum_x \sum_y p_{X,Y}(x,y) \log_2 p_{X,Y}(x,y) \\ &= -\sum_x \sum_y p_{X,Y}(x,y) \log_2 \big(p_Y(y) p_{X|Y}(x|y)\big) \\ &= -\sum_x \sum_y p_{X,Y}(x,y) \log_2 p_Y(y) - \sum_x \sum_y p_{X,Y}(x,y) \log_2 p_{X|Y}(x|y) \\ &= -\sum_y \log_2 p_Y(y) \sum_x p_{X,Y}(x,y) + H(X|Y) \\ &= -\sum_y \log_2 p_Y(y) \cdot p_Y(y) + H(X|Y) \\ &= H(Y) + H(X|Y). \end{aligned}$$

Второто равенство се доказва по подобен начин. Останалата част от теоремата следва от очевидните неравенства

$$H(Y|X) \leq H(Y), \quad H(X|Y) \leq H(X).$$

□

Забележка 3.8. Бихме могли да докажем равенството и като пресметнем математическото очакване от двете страни на равенството

$$\log p(x,y) = \log p(x) + \log p(y|x).$$

Следствие 3.9. Нека X и Y са независими случайни величини. Тогава

(*a*) $H(X,Y) = H(X) + H(Y)$;

(*b*) $H(X|Y) = H(X)$;

(*c*) $H(Y|X) = H(Y)$.

Доказателство. Аналогично на Теорема 3.7, като се използва равенството $p_{X,Y}(x,y) = p_X(x)p_Y(y)$. □

Следствие 3.10. За случайните величини X, Y и Z е всила равенството

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z).$$

Пример 3.11. Нека X и Y са случайни величини със следното съвместно разпределение:

X / Y	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Маргиналните разпределения на X и Y са съответно $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$ и $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Следователно $H(X) = \frac{7}{4}$ бита и $H(Y) = 2$ бита. Освен това $H(X|Y) = \frac{11}{8}$ бита, $H(Y|X) = \frac{13}{8}$ бита и $H(X,Y) = \frac{27}{8}$ бита. Ще отбележим, че $H(Y|X) \neq H(X|Y)$. Въпреки това $H(X) - H(X|Y) = H(Y) - H(Y|X)$, както твърдим в Теорема 3.7. □

Нека $I_{X,Y}(x,y)$ означава количеството информация, което $Y = y$ дава относно реализирането на $X = x$; то е равно на количеството информация, което дава реализирането на $X = x$ минус количеството, получено от $X = x$, при положение че е известно $Y = y$:

$$\begin{aligned} I_{X,Y}(x,y) &= -\log_2 p_X(x) + \log_2 p_{X|Y}(x|y) \\ &= -\log_2 \frac{p_X(x)}{p_{X|Y}(x|y)} \\ &= -\log_2 \frac{p_X(x)p_Y(y)}{p_{X,Y}(x,y)} \\ &= I_{Y,X}(y,x). \end{aligned}$$

Дефиниция 3.12. *Взаимна инфромация* $I(X;Y)$ на случайните величини X и Y наричаме очакваната стойност на $I_{X,Y}(x,y)$, т.е.

$$\begin{aligned} I(X;Y) &= \sum_x \sum_y p_{X,Y}(x,y) I_{X,Y}(x,y) \\ &= -\sum_x \sum_y p_{X,Y}(x,y) \cdot \log_2 \frac{p_X(x)p_Y(y)}{p_{X,Y}(x,y)} \\ &= -\sum_x \sum_y p_{X,Y}(x,y) \cdot \log_2 p_X(x)p_{X|Y}(x|y). \end{aligned}$$

$I(X;Y)$ се интерпретира като очакваното коичество информация, което Y дава относно X (или очакваното количество информация, което X дава относно Y). Тя измерва намаляването на несигурността за една случайна величина, дължащо се на знанието за друга случайна величина.

Теорема 3.13. Нека X и Y са дискретни случайни величини. Тогава

(*a*) $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$;

(*b*) $I(X;Y) = H(X) + H(Y) - H(X,Y)$;

(*c*) $I(X;X) = H(X)$;

(*d*) $I(X;Y) = I(Y;X)$.

Доказателство. (*a*) Можем да препишем дефиницията на взаимна информация във

вида:

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \\ &= \sum_{x,y} p_{X,Y}(x,y) \log \frac{p_{X|Y}(x|y)}{p_X(x)} \\ &= -\sum_{x,y} p_{X,Y}(x,y) \log p_X(x) + \sum_{x,y} p_{X,Y}(x,y) \log p_{X|Y}(x|y) \\ &= -\sum_{x} p_X(x) \log p_X(x) - \left(-\sum_{x,y} p_{X,Y}(x,y) \log p_{X|Y}(x|y) \right) \\ &= H(X) - H(X|Y). \end{aligned}$$

От съображения за симетрия, $I(X;Y) = H(Y) - H(Y|X)$.[4] От $H(X,Y) = H(X) + H(Y|X)$ получаваме (b). Накрая,

$$I(X;X) = H(X) - H(X|X) = H(X).$$

□

Понятието взаимна информация може да бъде въведено и чрез т.нар. разстояние на Kullback-Leibler [5].

Случайните величини от Пример 3.11 имат взаимна информация

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = 0.375 \text{ бита}.$$

[4] Това означава, че X носи толкова информация за Y, колкото и Y за X.

[5] The *relative entropy* (or *Kullback-Leibler distance*) between the probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) log \frac{p(x)}{q(x)} = E_p \left(\log \frac{p(X)}{q(X)} \right).$$

Here we use the convention $0 \log \frac{0}{q} = 0$, $p \log \frac{p}{0} = \infty$. The relative entropy is always nonnegative and is 0 iff $p(x) = q(x)$ for every x. This is not a true distance since it is not symmetric and does not satisfy the triangle inequality. Nevertheless, it is useful to think of relative entropy as of distance between distributions.

For example, if we know the true distribution $p(x)$ of a random variable, then we could construct a code with average description length $H(p)$. If, instead, we used a code for a distribution $q(x)$, we would need $H(p)+D(p \parallel q)$ bits on the average to describe the random variable.

Let $\mathcal{X} = \{0, 1\}$ and consider the two distributions p and q on $\mathcal{X}$: $p(0) = 1-r$, $p(1) = r$, $q(0) = 1-s$, $q(1) = s$. Then we have

$$\begin{aligned} D(p \parallel q) &= (1-r) \log \frac{1-r}{1-s} + r \log \frac{r}{s}, \\ D(q \parallel p) &= (1-s) \log \frac{1-s}{1-r} + s \log \frac{s}{r}. \end{aligned}$$

If $r = s$, then $D(p \parallel q) = D(q \parallel p) = 0$. If $r = 1/2$ and $s = 1/4$, then we can calculate

$$D(p \parallel q) = 1 - \frac{1}{2} \log 3, \; D(q \parallel p) = \frac{3}{4} \log 3 - 1.$$

This illustartes that in general $D(p \parallel q) \neq D(q \parallel p)$.

Let X and Y be random variables with a joint probability mass function $p_{X,Y}(x,y)$ and marginal probability mass functions $p_X(x)$ and $p_Y(y)$, respectively. The *mutual information* $I(X;Y)$ is defined as the relative entropy

Пример 3.15. (двоичен симетричен канал)
Дадена е случайна величина X, която приема стойност $X = 0$ или $X = 1$ с вероятност $\frac{1}{2}$. Можем да считаме, че този символ се предава по някакъв канал. Получателят приема стойност, която може да се различава от изпратената. Формално той наблюдава случайна величина Y, за която е известно, че $Y = X$ с вероятност $1 - p$ и $Y = 1 - X$ с вероятност p. Разпределението на Y може да се получи от това на X:

$$p_Y(0) = p_{Y|X}(0|0)p_X(0) + p_{Y|X}(0|1)p_X(1) = (1-p)\frac{1}{2} + p \cdot \frac{1}{2} = \frac{1}{2};$$

по подобен начин $p_Y(1) = \frac{1}{2}$. Имаме също така

$$p_{X,Y}(0,0) = p_{X,Y}(1,1) = \frac{1-p}{2}, \quad p_{X,Y}(0,1) = p_{X,Y}(1,0) = \frac{p}{2}.$$

Следователно

$$\begin{aligned} I(X;Y) &= -2\left\{\frac{1-p}{2} \cdot \log_2 \frac{\frac{1}{2}}{1-p} + \frac{p}{2} \cdot \log_2 \frac{\frac{1}{2}}{p}\right\} \\ &= 1 + p\log_2 p + (1-p)\log_2(1-p) \\ &= 1 - H(p). \end{aligned}$$

Това означава, че получателят разполага с $1 - H(p)$ бита информация за X от всеки приет символ Y. В случая, когато $p = \frac{1}{2}$, той не получава информация за X. Основна задача на теория на кодирането е създаване на такива начини за кодиране на данните, че получателят да може да извлече $1 - H(p)$ бита информация от всеки приет символ. □

Нека е дадена криптосистема $\mathfrak{S} = (\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$. Върху множеството от ключовете е дефинирана случайна величина K с разпределение $Pr_K\{K = k\}$. Открития текст и криптотекста можем да разглеждаме като редици от случайни величини, означени съответно с

$$\begin{aligned} M^n &= (M_0, M_1, \ldots, M_{n-1}) \\ C^\nu &= (C_0, C_1, \ldots, C_{\nu-1}. \end{aligned}$$

Очевидно е изпълнено $C^\nu = E_K(M^n)$. Тъй като E_K е взаимноеднозначно изображение, то е изпълнено

$$H(M^n \mid K, C^\nu) = 0. \tag{3.6}$$

between the joint distribution $p_{X,Y}(x,y)$ and the product distribution $p_X(x)p_Y(y)$, i.e.

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y) \parallel p(x)p(y)) \\ &= E_{p(x,y)}\left(\frac{p(X,Y)}{p(X)p(Y)}\right). \end{aligned}$$

Теорема 3.14. Нека $p(x)$ и $q(x)$ са разпределения върху множеството $\mathcal{X}$. Тогава $D(p \parallel q) \geq 0$, като равенство се достига тогава и само тогава, когато $p(x) = q(x)$ за всяко $x \in \mathcal{X}$.

Това означава, че ако са известни криптотекста и използвания ключ, то дешифрирането е еднозначно. Сигурността на всяка криптосистема е свързана с количеството информация за M^n, съдържащо се в C^ν. Колкото последното е по-малко, толкова системата е по-сигурна. В сила е следната оценка за взаимната информация на случайните величини M^n и C^ν.

Теорема 3.16. $I(M^n; C^\nu) \geq H(M^n) - H(K)$.

Доказателство. От (3.6) и Теорема 3.7 имаме

$$\begin{aligned} H(K|C^\nu) &= H(K|C^\nu) + H(M^n|K, C^\nu) = H(M^n, K|C^\nu) \\ &= H(M^n|C^\nu) + H(K|M^n, C^\nu) \geq H(M^n|C^\nu, \end{aligned}$$

т.е. при задаен криптотекст нашата несигурност за ключа е поне толкова голяма, колкото нашата несигурност за открития текст. Следователно

$$H(M^n|C^\nu) \leq H(K|C^\nu) \leq H(K),$$

откъдето

$$I(M^n; C^\nu) = H(M^n) - H(M^n|C^\nu) \geq H(M^n) - H(K).$$

□

Дефиниция 3.17. Една криптосистема наричаме *безусловно сигурна* (или казваме, че притежава *съвършена секретност*) ако за всеки две естествени числа n и ν взаимната информация на открития текст и криптотекста, разглеждани като случайни величини, е 0, т.е. $I(M^n; C^\nu) = 0$.

Следствие 3.18. Неравенството $H(M^n) \leq H(K)$ е необходимо условие за безусловната сигурност на всяка криптосистема.

Да отбележим, че $I(M^n; C^\nu) = 0$ тогава и само тогава, когато $H(M^n) = H(M^n|C^\nu)$. Това означава, че несигурността ни за открития текст не намалява от това, че наблюдаваме криптотекста.

Възможна е друга дефиниция на понятието безусловна сигурност, която е еквивалентна на дадената по-горе. Нека $M = \{m_1, \ldots, m_s\}$ е множеството на откритите текстове, $C = \{c_1, \ldots, c_u\}$ – множеството на криптотекстовете, а $E = \{E_1, \ldots, E_k\}$ – множеството на всички шифриращи трансформации. Нека $p(m_i)$ е априорната вероятност за изпращане на съобщението m_i, а с $p_j(m_i)$ – вероятността да е изпратено съобщението m_i, ако е известно, че е получен криптотекста c_j.

Дефиниция 3.19. Една криптосистема на наричаме безусловно сигурна ако завсяко $1 \leq i \leq s$ и всяко $1 \leq j \leq t$ е в сила $p_j(m_i) = p(m_i)$.

Забележка 3.20. Дефиниции 3.17 и 3.19 са еквивалентни.

Нека най-напред да предположим, че $H(M^n) = H(M^n|C^\nu)$. Тогава

$$\begin{aligned}\sum p_M(M^n)\log_2 p_M(M^n) &= \sum_{M^n}\sum_{C^\nu} p_C(C^\nu)p_{M|C}(M^n|C^\nu)\log_2 p_{M|C}(M^n|C^\nu)\\ &\leq \sum_{M^n}\sum_{C^\nu} p_C(C^\nu)p_M(M^n)\log_2 p_M(M^n)\\ &= \sum_{M^n} p_M(M^n)\log_2 p_M(M^n)\sum_{C^\nu} p_C(C^\nu)\\ &= \sum_{M^n} p_M(M^n)\log_2 p_M(M^n).\end{aligned}$$

Равенство се достига тогава и само тогава, когато за всяко съобщение с дължина n и всеки криптотекст с дължина ν е в сила $p_{M|C}(M^n|C^\nu) = p_M(M^n)$.

Сега да допуснем, че е в сила Дефиниция 3.19, т.е. $p(m_i) = p_M(M^n) = p_j(m_i) = p_{M|C}(M^n|C^\nu)$. Тогава имаме

$$\begin{aligned}H(M^n|C^\nu) &= -\sum_{M^n}\sum_{C^\nu} p_C(C^\nu)p_{M|C}(M^n|C^\nu)\log_2 p_{M|C}(M^n|C^\nu)\\ &= -\sum_{M^n}\sum_{C^\nu} p_C(C^\nu)p_M(M^n)\log_2 p_M(M^n)\\ &= -\sum_{M^n} p_M(M^n)\log_2 p_M(M^n)\\ &= H(M^n).\end{aligned}$$

Възможна е и трета дефиниция за безусловно сигурна криптосистема, която е еквивалентна на първите две.

Дефиниция 3.21. В означенията от Дефиниция 3.19 нека $p(c_j) = p_C(C^\nu)$ е априорната вероятност за получаване на криптотекста $c_j = C^\nu$ и нека $p_i(c_j) = p_{C|M}(C^\nu|M^n)$ е условната вероятност за получаване на c_j, при условие че е изпратен откритият текст $m_i = M^n$. Една криптосистема наричаме безусловно сигурна. ако $p_i(c_j) = p(c_j)$.

Лемата по-долу следва непосредствено от правилото на Бейс.

Лема 3.22. $p_j(m_i)p(c_j) = p_i(c_j)p(m_i)$.

От Лемма 3.22 лесно следва еквивалентността на Дефиниции 3.19 и 3.21.

3.3 Ентропия на езика

Да разгледаме криптосистема $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ и нека крипотекстът $y_1y_2\ldots y_n \in \mathcal{C}^n$ е получен от открития текст $x_1x_2\ldots x_n \in \mathcal{P}^n$ при използване на един и същи ключ. Опонентът O притежава неограничени изчислителни възможности и провежда атака при известен криптотекст. Приемаме, че откритият текст е на някакъв естествен език, например английски, и това е известно на опонента. В общия случай O може да отхвърли много ключове като невъзможни, но остават и много възможни ключове,

един от които е правилният. останалите възможни, но некоректни ключове, наричаме *неистински* (spurious keys). Например при шифър на Цезар криптотекстът `WNAJW` може да бъде получен от два смислени открити текста: `arena` (при ключ $k = 22$) и `river` (при ключ $k = 5$). Единият от тях е правилен, а другият – неистински, но опонентът няма как да ги различи, дори да притежава безкрайни изчислителни възможности. Нашата цел е да дадем граница за броя на неистинските ключове.

Най-напред ще дефинираме понятието *ентропия на буква* за естествения език L и ще я означаваме с H_L. Тя е мярка за средното количество информация, съдържащо се в буква от низ, представляващ открит текст на английски език. Случаен низ от букви има $\log_2 26 \approx 4.76$ бита ентропия (на буква). Като апроксимация от първи ред можем да вземем ентропията на случайна величина $P = (\mathcal{X}, p)$, където $\mathcal{X} = \{A, \ldots, Z\}$, а $p_X(\alpha)$ е вероятността за появяване на $\alpha \in \mathcal{X}$ в английски текст. В този случай $H(p) \approx 4.19$. Разбира се, две последователни наблюдения на съседни букви в английски текст не са независими събития. Например в английския език след буквата `q` почти винаги следва `u`. За апроксимация от втори ред можем да използваме вероятностите на всички двойки букви (биграми), което дава $H(P^2) \approx 3.90$.

В общия случай дефинираме случайна величина P^n с вероятностно разпределение това на всички n-грами в текста. така стигаме до следната дефиниция. Нека L е естествен език. Ентропия на L наричаме величината

$$H_L = \lim_{n \to \infty} \frac{H(P^n)}{n},$$

а излишък на L е величината

$$R_L = 1 - \frac{H_L}{\log_2 |\mathcal{P}|}.$$

Различни експерименти дават следните граници за H_L:

$$1 \leq H_L \leq 1.5$$

В [18] е дадена оценката $H_l \approx 1.34$. Така средното съдържание на информация в буква от английски текст е между един бит и един и половина бита.

При зададени вероятностни разпределения върху $\mathcal{K}$ и $\mathcal{P}^n$ ще дефинираме индуцираното разпределение върху $\mathcal{C}^n$ - множеството от n-грамите криптотекст (както направихме в предния раздел за $n = 1$). $\mathcal{P}^n$ е случайна величина, представяща n-грамите открит текст, а $\mathcal{C}^n$ е случайна величина, представяща n-грамите криптотекст. За $y \in \mathcal{C}^n$ дефинираме

$$K(y) = \{k \in \mathcal{K} \mid \exists x \in \mathcal{P}^n, p_{P^n}(x) > 0, E_k(x) = y\}.$$

Това са всички ключове, шифриращи в y текстове, които се появяват с положителна вероятност, или, с други думи, множеството от ключове, довеждащи смислен низ открит текст в y. Ако y е наблюдаваният криптотекст, то броят на неизтинските ключове е $k(y) - 1$. Средният брой неистински ключове по всички възможни низове

с дължина n означаваме с $\overline{s}_n$.За тази величина имаме:

$$\begin{aligned}\overline{s}_n &= \sum_{y\in C^n} p(y)(|K(y)|-1)\\ &= \sum_{y\in C^n} p(y)|K(y)| - \sum_{y\in C^n} p(y)\\ &= \sum_{y\in C^n} p(y)|K(y)| - 1.\end{aligned}$$

Известно е(?), че

$$H(K|C^n) = H(K) + H(P^n) - H(C^n).$$

Можем да приемем, че

$$H(P^n) \approx nH_L = (n(1-R_L)\log_2|\mathcal{P}|$$

за достатъчно големи n. Освен ова е ясно, че

$$H(C^n) \leq n\log_2|\mathcal{C}|.$$

Ако $|\mathcal{P}| = |\mathcal{C}|$, получаваме, че

$$H(K|C^n) \geq H(K) - nR_L\log_2|\mathcal{P}|. \tag{3.7}$$

Сега ще свържем условната ентропия $H(K|C^n)$ с броя на неистинските ключове s_n. Използвайки неравенството на Йенсен за $f(x) = \log_2 x$, получаваме оценката:

$$\begin{aligned}H(K|C^n) &= \sum_{y\in C^n} p(y)H(K|y)\\ &\leq \sum_{y\in C^n} p(y)\log_2|K(y)|\\ &\leq \log_2 \sum_{y\in C^n} p(y)|K(y)|\\ &= \log_2(\overline{s}_n+1).\end{aligned}$$

Комбинирайки полученото неравенство с (3.7, стигаме до

$$\log_2(\overline{s}_n+1) \geq H(K) - nR_L\log_2|\mathcal{P}|.$$

В случая, когато ключовете са равновероятни имаме следния резултат.

Теорема 3.23. *Нека $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ е криптосистема, за която $|\mathcal{P}| = |\mathcal{C}|$ и за която ключовете са равновероятни. Нека R_L е излишъкът на съответния език.Тогава за зададен криптотекст с дължина n, където n е достатъчно голямо, очакваният брой на неистинските ключове е равен на*

$$\overline{s}_n \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1.$$

Величината $|\mathcal{K}||\mathcal{P}|^{nR_L} - 1$ се приближава много бързо към 0, когато n расте неограничено. Тази формула може и да не е точна за малки n поради това, че в тези случаи $H(P^n)/n$ не е добра оценка за H_L.

Дефиниция 3.24. *Точка на единственост* за криптосистема се дефинира като стойността на n, за която очакваният брой неистински ключове става нула. Точката на единственост означаваме с n_0. Това е необходимото количество криптотекст, необходимо на опонента (който разполага с неограничен изчислителен ресурс и достатъчно време) за да може еднозначно да определи ключа.

Ако положим в горната теорема $\overline{s}_n = 0$ и решим относно n, получаваме

$$n_0 \approx \frac{\log_2 |\mathcal{K}|}{R_L \log_2 |\mathcal{P}|}.$$

да разгледаме простата субституция. В тази система имаме $|\mathcal{P}| = 26$, $|\mathcal{K}| = 26!$. Ако приемем, че $R+L = 0.75$, то получаваме следната оценка за точката на единственост:

$$n_o \approx \frac{88.4}{(0.75 \times 4.7 \approx 25}.$$

Това отразява факта, че при задаен криптотекст с дължина 25 обикновено е възможно единствено дешифриране.

3.4 Оценяване на сигурността на криптосистема

Сигурността на криптографските примитиви може да се оценява в различни модели при различни допускания за възможностите на опонента. По-долу излагаме една възможна класификация, която в не може да се счита за изчерпваща.

(1) *Безусловна сигурност* (unconditional security). Това е най-високото ниво на сигурност, която една криптосистема може да притежава. При нея се предполага, че опонентът разполага с неограничен изчислителен ресурс и единственият въпрос е дали наличната информация е достатъчна заразбиване на шифъра. В този случай несигурността за открития текст трябва да съвпада с априорната несигурност за открития текст, т.е. криптотекстът не трябва да дава никаква информация за открития текст. Може да се докаже, че еднократният ключ е безусловно сигурен шифър. В общия случай криптосистемите не са безусловно сигурни. Асиметричните системи по правило не са безусловно сигурни.

(2) *Сигурност по сложност* (complexity theoretic security). В този случай се дефинира изчислителен модел, при който се предполага, че опонентът разполага с определени изчислителни възможности (в термините на теория на алгоритмите). Най-често се предполага, че те са ограничени от полиномиални алгоритми. Анализът на асимптотичното поведение на алгоритмите се извършва за най-лошия случай и в много от случаите няма голяма практическа стойност (вж. напр. криптосистемата на Merkle-Hellman, основана на задача за раницата), но все пак води до добро общо разбиране за сигурността на системите.

(3) *Доказуема сигурност* (provable security). Една криптосистема е доказуемо сигурна, ако разбиването ѝ е еквивалентно на решаването на добре известна задача, за която се предполага, че е трудна. Обикновено това е теоретико-числова задача като, например, разлагане на цяло число на прости множители или намиране на дискретен логаритъм. Тук "доказуемо сигурна" означава "доказуемо сигурна при определени допускания". Доказуемата сигурност може да се разглежда като специален случай на изчислителната сигурност.

(4) *Изчислителна сигурност* (computational security). Изчислителната сигурност почива на оценка за обема изчисления, необходими на известните към момента методи за разбиване на дадена криптосистема. Предполага се, че системата е добре изследвана и е известно, кои са релевантните атаки.Една предполагаема криптографска техника се счита за изчислително сигурна, ако необходимите изчислителни ресурси за разбиването ѝ (използвайки най-добрите известни атаки) надхвърлят многократно изчислителните ресурси на хипотетичния опонент. Такава сигурност се нарича понякога практическа сигурност. Най-добрите известни системи с публичен ключ са в този клас.

(5) *Сигурност ad hoc* (ad hoc security). Тази сигурност е налична, ако съществуват убедителни аргументи, че всяка успешна атака изисква много по-големи ресурси (памет, време) от тези, с които разполага визирания опонент (можем да визираме слаби опоненти). Това е най-честият подход, когато разглеждаме криптографски протоколи. В същото време той е и най-неудовлетворителния. Твърдения за сигурност от този вид са често под въпрос и оставят заплаха от непредвидени атаки.